PSY 652 Lab: Module 3 Activity
September 18, 2019

The following exercises are intended to provide hands-on practice conceptualizing how different types of Questionable Research Practices (QRPs) impact study results and biases in published literature. While many types of QRPs exist, here we'll specifically focus on two types of HARKing: Cherry Picking and Question Trolling. We've included two tables from Murphy & Aguinis' 2019 article that used simulations to examine how Cherry Picking and Question Trolling impact statistical results. Answer the questions below using the tables and provided terms.

Note: *Without HARKing, the true parameter for the average correlation coefficient (mean r) across studies is .20 in these simulations.*

Definitions of key terms:

1.) ***QRPs:*** *Questionable Research Practices*
2.) ***HARKing:*** *Hypothesizing After Results are Known*
3.) ***Forms of HARKing:***
    a.) Cherry Picking: Involves searching through data that includes alternative measures or samples to find the results that offer the strongest possible support for a particular hypothesis or research question a study was designed to investigate. *Tends to be associated with a more homogenous set of effects in the published literature*.
    b.) Question Trolling: Involves searching through data that includes several different constructs, interventions, or relationships to find seemingly notable results. *Tends to be associated with more heterogeneous sets of effects in the published literature.*
4.) ***Sample size:*** The number of participants included in a study.
5.) ***Pool size:*** The total number of statistical tests a research team conducts and chooses from when selecting results to publish (e.g., if a researcher conducts 10 tests and elects to only publish the one test that detected the largest effect size, pool size = 10; if a researcher conducts 2 tests and elects to only publish the test that detected a larger effect size, pool size = 2).
6.) ***QRP Prevalence:*** The percentage of studies that employed a specific type of QRP. In the studies included below, QRP prevalence was specified as a simulation parameter.
7.) ***Mean r:*** The average correlation coefficient reported across a set of studies. In this case, mean r is our proxy for effect size, where larger r values reflect larger effect sizes.
8.) ***Question Trolling Heterogeneity* (aka effect size SD)**: In this context, effect size SD represents the upper bound for variability in effect sizes in each set of studies. The more haphazardly a research team trolls for questions (i.e., the more variable relationships they examine), the more variability they will observe in effect sizes (i.e., a wider range of effect sizes to choose from when selecting results to publish).

Section 1: Impact of cherry picking results

Below are results from Murphy & Aguinis' (2019) HARKing simulation.

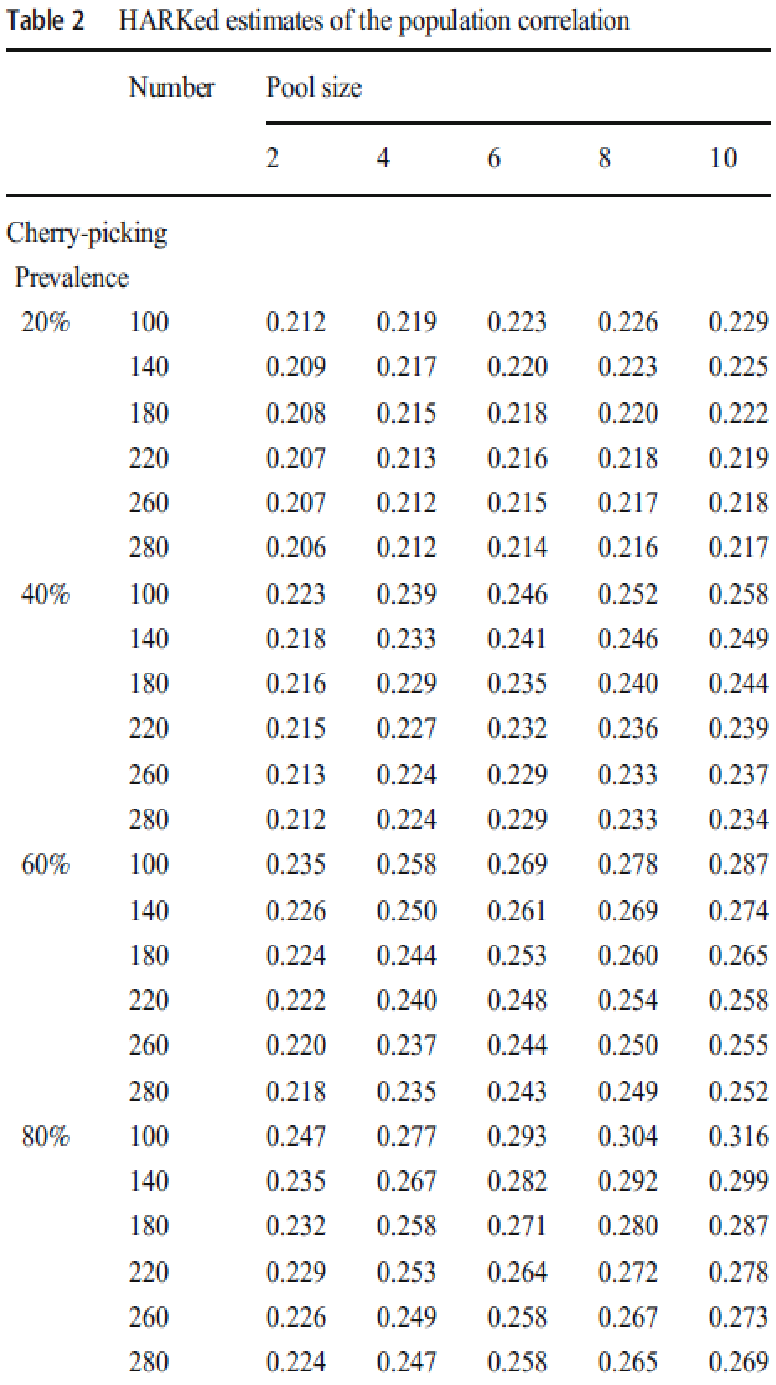


Table 2 HARKed estimates of the population correlation

	Number	Pool size				
		2	4	6	8	10
Cherry-picking Prevalence						
20%	100	0.212	0.219	0.223	0.226	0.229
	140	0.209	0.217	0.220	0.223	0.225
	180	0.208	0.215	0.218	0.220	0.222
	220	0.207	0.213	0.216	0.218	0.219
	260	0.207	0.212	0.215	0.217	0.218
	280	0.206	0.212	0.214	0.216	0.217
40%	100	0.223	0.239	0.246	0.252	0.258
	140	0.218	0.233	0.241	0.246	0.249
	180	0.216	0.229	0.235	0.240	0.244
	220	0.215	0.227	0.232	0.236	0.239
	260	0.213	0.224	0.229	0.233	0.237
	280	0.212	0.224	0.229	0.233	0.234
60%	100	0.235	0.258	0.269	0.278	0.287
	140	0.226	0.250	0.261	0.269	0.274
	180	0.224	0.244	0.253	0.260	0.265
	220	0.222	0.240	0.248	0.254	0.258
	260	0.220	0.237	0.244	0.250	0.255
	280	0.218	0.235	0.243	0.249	0.252
80%	100	0.247	0.277	0.293	0.304	0.316
	140	0.235	0.267	0.282	0.292	0.299
	180	0.232	0.258	0.271	0.280	0.287
	220	0.229	0.253	0.264	0.272	0.278
	260	0.226	0.249	0.258	0.267	0.273
	280	0.224	0.247	0.258	0.265	0.269

1) Conceptually, what does this table demonstrate?

a) Why does mean r tend to increase with increasing pool size?

Answer: The larger the pool size, the more tests that were run in order to obtain the reported effect size. Running more tests increases probability of detecting larger effect sizes. Thus, mean effect size increases with increasing pool size.

b) Why does mean r tend to decrease with increasing sample size?

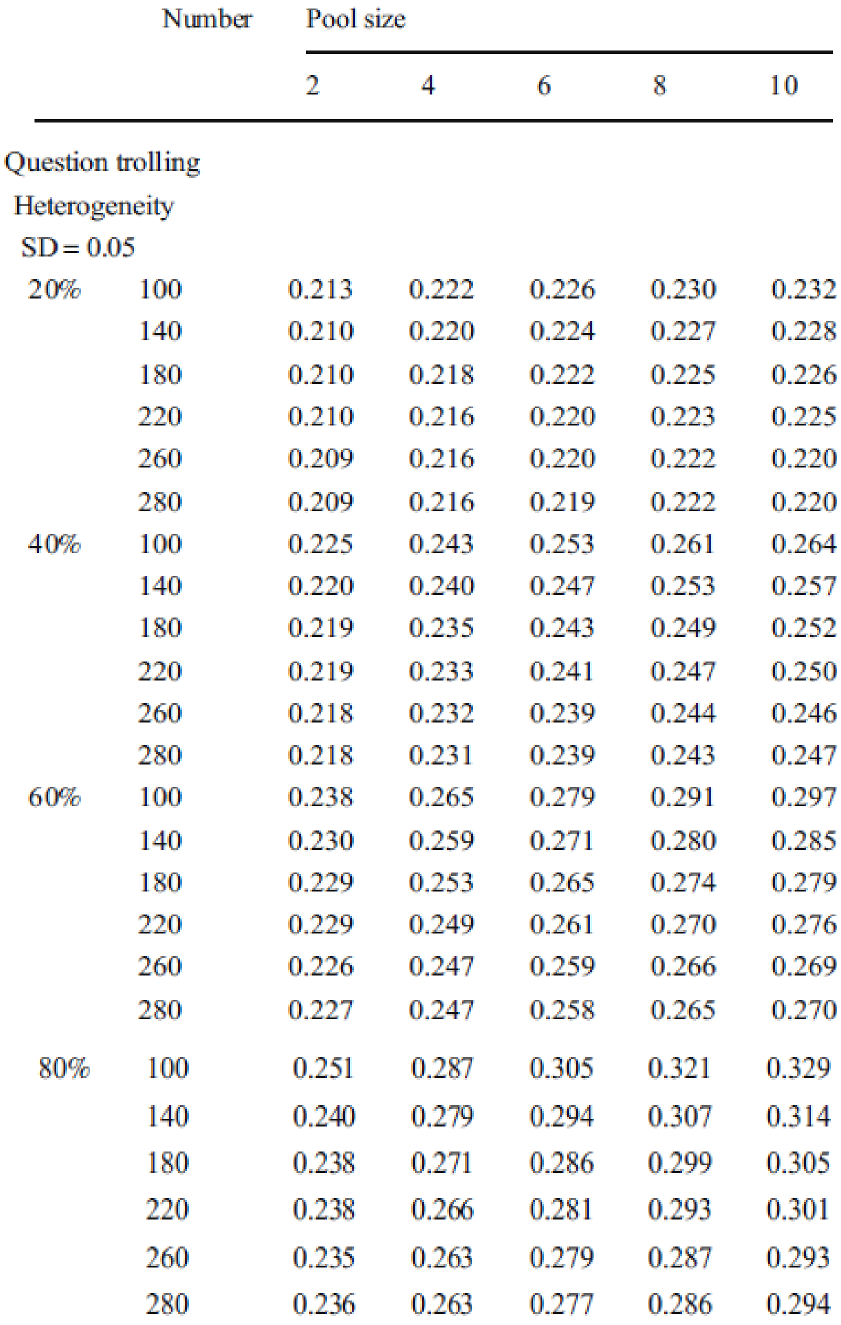
Answer: As sample size increases, more power is established and there is a higher probability of detecting smaller effects, if they exist, in the data. Therefore, smaller effects are more prevalent among studies that utilized larger sample sizes. Given the frequent use of small sample sizes in behavioral and social sciences research, there tends to be a bias in the literature towards large effect sizes.

c) What is the impact of increasing cherry picking prevalence?

Answer: As the prevalence of studies utilizing cherry picking increases, the average correlation effect size increases. This is a direct result of more studies selecting the most inflated result because of biases towards larger effects in published literature.

Section 2: Question trolling

Below are results from Murphy & Aguinis' (2019) HARKing simulation.



	Number	Pool size				
		2	4	6	8	10
Question trolling						
Heterogeneity						
SD = 0.05						
20%	100	0.213	0.222	0.226	0.230	0.232
	140	0.210	0.220	0.224	0.227	0.228
	180	0.210	0.218	0.222	0.225	0.226
	220	0.210	0.216	0.220	0.223	0.225
	260	0.209	0.216	0.220	0.222	0.220
	280	0.209	0.216	0.219	0.222	0.220
40%	100	0.225	0.243	0.253	0.261	0.264
	140	0.220	0.240	0.247	0.253	0.257
	180	0.219	0.235	0.243	0.249	0.252
	220	0.219	0.233	0.241	0.247	0.250
	260	0.218	0.232	0.239	0.244	0.246
	280	0.218	0.231	0.239	0.243	0.247
60%	100	0.238	0.265	0.279	0.291	0.297
	140	0.230	0.259	0.271	0.280	0.285
	180	0.229	0.253	0.265	0.274	0.279
	220	0.229	0.249	0.261	0.270	0.276
	260	0.226	0.247	0.259	0.266	0.269
	280	0.227	0.247	0.258	0.265	0.270
80%	100	0.251	0.287	0.305	0.321	0.329
	140	0.240	0.279	0.294	0.307	0.314
	180	0.238	0.271	0.286	0.299	0.305
	220	0.238	0.266	0.281	0.293	0.301
	260	0.235	0.263	0.279	0.287	0.293
	280	0.236	0.263	0.277	0.286	0.294

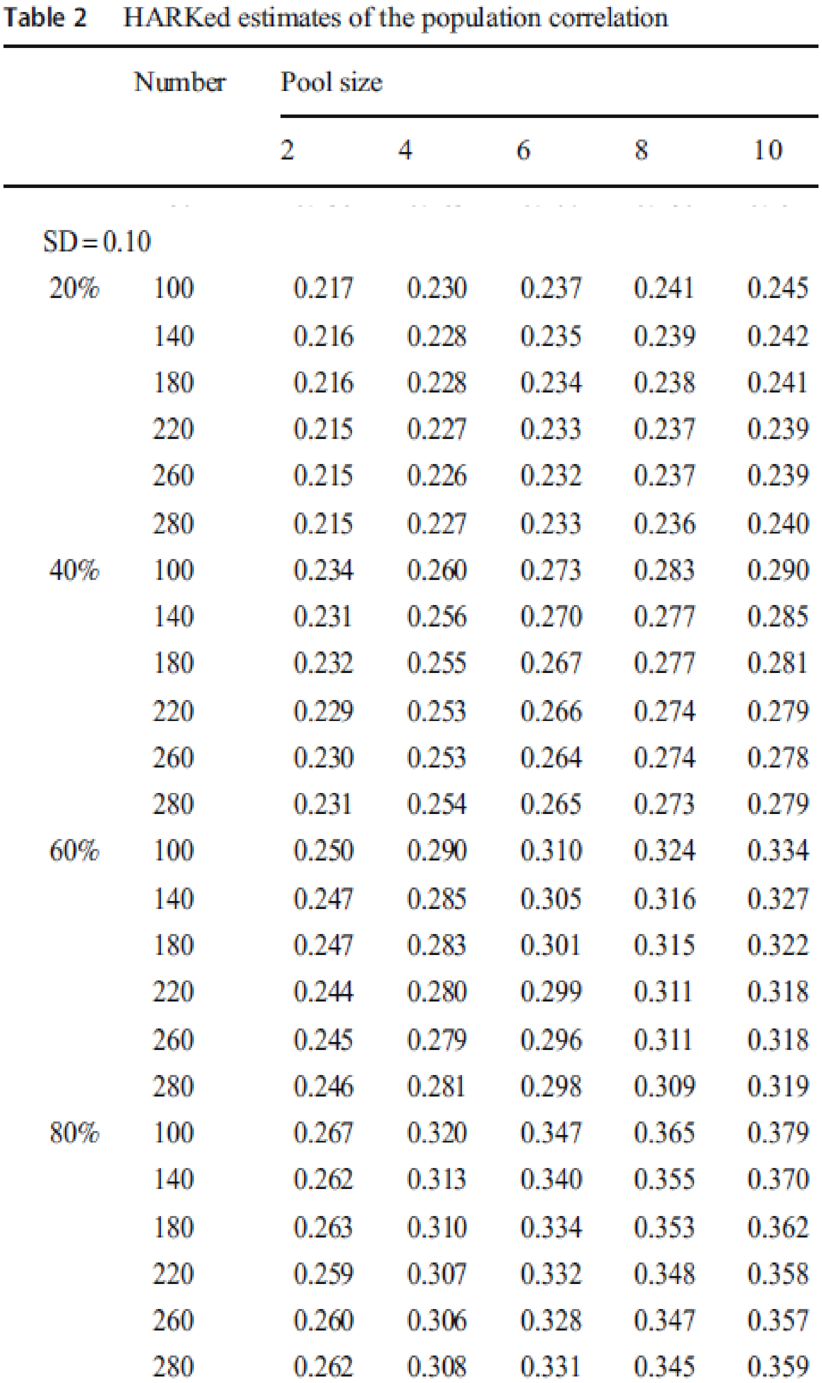


Table 2 HARKed estimates of the population correlation

	Number	Pool size				
		2	4	6	8	10
SD = 0.10						
20%	100	0.217	0.230	0.237	0.241	0.245
	140	0.216	0.228	0.235	0.239	0.242
	180	0.216	0.228	0.234	0.238	0.241
	220	0.215	0.227	0.233	0.237	0.239
	260	0.215	0.226	0.232	0.237	0.239
	280	0.215	0.227	0.233	0.236	0.240
40%	100	0.234	0.260	0.273	0.283	0.290
	140	0.231	0.256	0.270	0.277	0.285
	180	0.232	0.255	0.267	0.277	0.281
	220	0.229	0.253	0.266	0.274	0.279
	260	0.230	0.253	0.264	0.274	0.278
	280	0.231	0.254	0.265	0.273	0.279
60%	100	0.250	0.290	0.310	0.324	0.334
	140	0.247	0.285	0.305	0.316	0.327
	180	0.247	0.283	0.301	0.315	0.322
	220	0.244	0.280	0.299	0.311	0.318
	260	0.245	0.279	0.296	0.311	0.318
	280	0.246	0.281	0.298	0.309	0.319
80%	100	0.267	0.320	0.347	0.365	0.379
	140	0.262	0.313	0.340	0.355	0.370
	180	0.263	0.310	0.334	0.353	0.362
	220	0.259	0.307	0.332	0.348	0.358
	260	0.260	0.306	0.328	0.347	0.357
	280	0.262	0.308	0.331	0.345	0.359

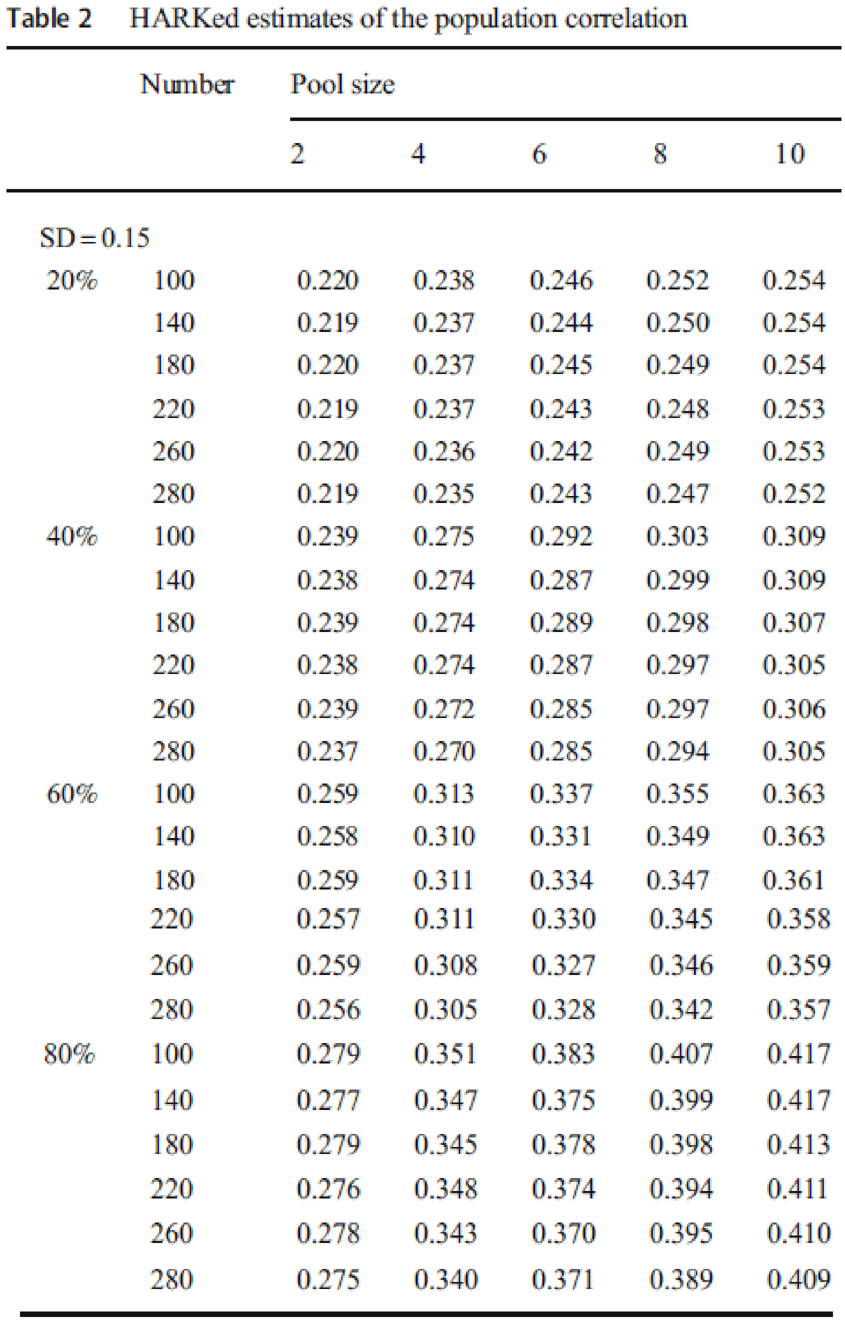


Table 2 HARKed estimates of the population correlation

	Number	Pool size				
		2	4	6	8	10
SD = 0.15						
20%	100	0.220	0.238	0.246	0.252	0.254
	140	0.219	0.237	0.244	0.250	0.254
	180	0.220	0.237	0.245	0.249	0.254
	220	0.219	0.237	0.243	0.248	0.253
	260	0.220	0.236	0.242	0.249	0.253
	280	0.219	0.235	0.243	0.247	0.252
40%	100	0.239	0.275	0.292	0.303	0.309
	140	0.238	0.274	0.287	0.299	0.309
	180	0.239	0.274	0.289	0.298	0.307
	220	0.238	0.274	0.287	0.297	0.305
	260	0.239	0.272	0.285	0.297	0.306
	280	0.237	0.270	0.285	0.294	0.305
60%	100	0.259	0.313	0.337	0.355	0.363
	140	0.258	0.310	0.331	0.349	0.363
	180	0.259	0.311	0.334	0.347	0.361
	220	0.257	0.311	0.330	0.345	0.358
	260	0.259	0.308	0.327	0.346	0.359
	280	0.256	0.305	0.328	0.342	0.357
80%	100	0.279	0.351	0.383	0.407	0.417
	140	0.277	0.347	0.375	0.399	0.417
	180	0.279	0.345	0.378	0.398	0.413
	220	0.276	0.348	0.374	0.394	0.411
	260	0.278	0.343	0.370	0.395	0.410
	280	0.275	0.340	0.371	0.389	0.409

2. Conceptually, what does this table demonstrate?

a. Why does mean r tend to increase with increasing pool size?

Answer: Again, the larger the pool size, the more tests that were run in order to obtain the reported effect size. Running more tests increases probability of detecting larger effect sizes. Thus, mean effect size increases with increasing pool size.

b. Why does mean r tend to decrease with increasing sample size?

Answer: Again, as sample size increases, more power is established and you have a higher probability of detecting small effects, if they exist, in the data. Therefore, smaller effects are more prevalent in studies that utilized larger sample sizes. Given the frequent use of small sample sizes in behavioral and social sciences research, there tends to be a bias in the literature towards large effect sizes.

c. What is the impact of increasing question trolling prevalence?

Answer: As the prevalence of studies utilizing question trolling practice increases, the average correlation effect size increases. This is a direct result of more studies choosing the most inflated result because of the desirability of larger effect sizes for publication.

d. Why is it useful to include Question Trolling Heterogeneity (effect size SD) in the question trolling simulation?

Answer: In these simulations, SD served as a proxy for how haphazardly researchers are engaging in question trolling. Put another way, Question Trolling Heterogeneity provides a way to estimate the degree to which the *extent* of question trolling impacts the range of effect sizes reported in studies that are investigating similar research questions.

For example, if researchers examine relationships between 500 variables to try to identify large effect sizes for publication (i.e., a very haphazard practice), they would likely identify a wide range of effect sizes to choose from (i.e., a larger SD for detected effect sizes). In contrast, if researchers only troll for significant results between 5 variables (i.e., a less haphazard approach), they would likely identify a smaller range of effect sizes (i.e., a smaller SD for detected effect sizes).

e. What happens as the Question Trolling Heterogeneity (effect size SD) increases?

Answer: The more haphazardly a research team trolls for questions (i.e., the more variable relationships they examine), the more variability they will observe in effect sizes (i.e., a wider range of effect sizes to choose from when selecting results to publish). Here, the mean effect size generally increases with increasing effect size SD because researchers were able to select from a wider range of effects, thereby increasing the probability of having larger effects included their pool of results to choose from.

3. What four factors affect the impact of bias produced from Cherry Picking & Question Trolling?

Answer:

1. Sample Size
2. Size of the pool of results the researcher is able to choose from
3. The heterogeneity of the population effects that underlie the pool of sample statistics
4. Prevalence of the HARKing procedure

4. Based on these tables, is cherry picking or question trolling more likely to be introducing biases into published literature? How did you reach your conclusion?

Answer: Although both forms of HARKing contribute to bias in the published literature, results from these simulations suggest that Cherry Picking introduces *less* bias than Question Trolling. As seen in the Cherry Picking simulation results table, the difference between the parameter correlation value and the obtained correlation value does not exceed .10 correlation units (for the pool of effect sizes evaluated here if the sample size is at least 140). However, the results from the Question Trolling simulation table show that the correlation effect size may differ by more than .20 from the true parameter value.

In Cherry Picking, the researcher knows what the research question is before developing a post hoc hypothesis, which generally results in a narrower range of effects to choose from (fewer variable relationships examined). Question Trolling generally involves conducting many tests using a heterogeneous set of measures. This leads to a wider range of potential research questions and results to choose from, causing a larger bias towards large effects in the literature.

5. What are three ways that researchers, editors, and journals can respond to Questionable Research Practices and the current “reproducibility crisis” in published literature?

Answer: Many potential solutions and current directions for best practices have been proposed. Here are some examples:

- Study pre-registration, in which journals can provisionally guarantee publication of results regardless of effect sizes and study significance.
- Requests for replication by journals (including development of clear criteria for what a quality replication would entail)
- Registered reports
- Make it common practice for journals to require power analyses, effect sizes, and effect size justifications to be included in published studies.
- Implement more meta-analyses to identify and examine potential biases in different areas of published literature (e.g., determine if results vary across studies asking equivalent research questions, and evaluate if question trolling or other factors may be at play).